OE5510 - Machine Learning for Ocean Engineers (Fall 2024)

Tutorial 01 - Dynamic Mode Decomposition

August 7, 2024

Dataset

Load the data set for fluid flow past a cylinder (you can download the dataset from the following link http://DMDbook.com). The data file can be found at the following location `DATA/DATA/FLUIDS/CYLINDER_ALL.mat`. Each column is a flow field that has been reshaped into a vector. Note that you have to use the variables `UALL` and `VALL` from the dataset for $[X_{vel}]$ and $[Y_{vel}]$ velocity fields respectively. The dataset contains 151 time snapshots which is the number of columns, and each column has 199×449 elements where 199 corresponds to number of rows and 449 correspond to number of columns in the 2D representation of the flow fields. In the image of the flow field, considering the origin at the bottom left corner and x and y axis along the horizontal (positive towards right) and vertical directions (positive upwards) respectively, the cylinder location is present at coordinate (0.5, 1) and the diameter of the cylinder is 1 unit.

Questions

Before starting the assignment questions try to visualize the data set by reshaping a few time snapshots from the matrices $[X_{vel}]$ and $[Y_{vel}]$. Create an animation from a series of sequential snapshots to visualize the velocities along x and y directions in the flow field. Plotly and Matplotlib animations offer an easy approach to generate dynamic plots in a notebook. However, you are free to try out any other approach to plot your data.

As a submission for the assignment you are expected to submit a **report** along with your **codes** in a **zip file**. Attempt the following questions using the above dataset and include the plots and results as mentioned for each question in your report.

1. Compute the SVD of this data set and plot the singular value spectrum for both $[X_{vel}]$ and $[Y_{vel}]$ data separately. Note your observations based on the SVD spectrum plots. The $[\hat{U}]$ from your SVD contains eigenflow fields and the $[\Sigma][V]^*$ represents the amplitudes of these eigenflows as the flow evolves in time. Plot the first three eigenflow fields as an animation.

2. Compute and mention the values of r (number of modes) needed to capture 90%, 99%, and 99.9% of the flow energy based on the singular value spectrum (recall that energy is given by the Frobenius norm squared). Write a code to plot the reconstructed movie for a chosen truncation value r. Plot the movies for each of three truncation values computed previously and compare the fidelity. Note your observations about the fidelity in the report along with appropriate images as

required. For each of the three cases also compute and report the squared Frobenius norm of the error between the true matrices $[X_{vel}]$, $[Y_{vel}]$ and the reconstructed matrices $[\tilde{X}_{vel}]$, $[\tilde{Y}_{vel}]$ respectively.

3. Fix a value ($r = 10$) and compute the truncated SVD. Each column $\{w_k\} \in \mathbb{R}^{10}$ of the matrix $[W] = [\tilde{\Sigma}][V]^*$ represents the mixture of the first 10 eigenflows in the k-th column of $[X]$. Verify this by comparing the k-th snapshot of $[X]$ with $[\tilde{U}]\{w_k\}$. For the report, choose $k = 89$ and show the plots for that column alone.

4. Now, using the same value of $r = 10$ we will build a linear regression model for how the amplitudes $\{w_k\}$ evolve in time. This will be a dynamical system of the form:

$$\{w_{k+1}\} = [A]\{w_k\}$$

   To obtain this dynamical system model, create a matrix $[W]$ with the first 1 through $m-1$ columns of $[\tilde{\Sigma}][V]^*$ and another matrix $[W]'$ with the 2 through m columns of $[\tilde{\Sigma}][V]^*$. We will now try to solve for a best-fit $[A]$ matrix so that

$$[W]' \approx [A][W]$$

   Compute the SVD of $[W]$ and use this to compute the pseudo-inverse of $[W]$ to solve for $[A]$. Compute the eigenvalues of $[A]$ and plot them in the complex plane.

5. Use this $[A]$ matrix to advance the state $\{w_k\} = [A]^{k-1}\{w_1\}$ starting from $\{w_1\}$. Plot the reconstructed flow field using these predicted amplitude vectors and compare with the true values. You can choose random snapshot of your choice to make this comparison or plot the movie of the flow fields as before.

   Bonus : You are encouraged to test with different values of m in the last 2 questions. At what value of m according to you the performance of this method starts to degrade? Could you relate to why does this happen at this value of m based on the SVD?